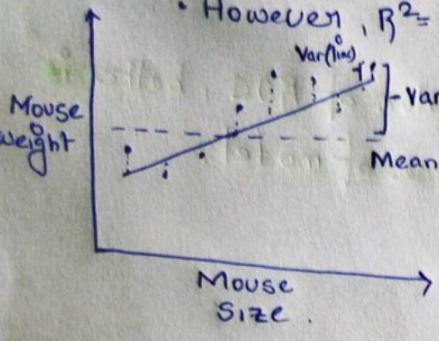
R, R^2 and Adjusted R^2

R → Correlation value is denoted by R.
- Correlation value close to 1 to -1 are good and tells you two quantitative variables (i.e. weight and size) are strongly related.

R^2 → In R^2 is very similar and interpretation is also easier.
- It not obvious that (regular) $R = 0.7$ is twice as good as $R = 0.35$
- However, $R^2 = 0.7$ is 2 times good as $R^2 = 0.35$.



$$R^2 = \frac{Var(mean) - Var(line)}{Var(mean)}$$

Suppose $R^2 = 0.81 = 81\%$.
There is 81% less variation around the line than the mean

OR

The size/weight relationship accounts for 81% of variation.

This means the most of the variation in the data is explained by the size/weight relationship.

Suppose $R^2 = 0.06 = 6\%$, There is only 6% less variation around the line than the mean or the weight/size relationship accounts for 6% of variation. So we have to find remaining 94%, why the variation is happening. In other word hardly any of the variation in the data is explained by the variable.

R^2 - Main Ideas.
- If someone gives you a value of plain old R, just square it.
- R^2 is the percentage of variation explained by relationship between two variables.

So which one is better $R = 0.7$ than $R = 0.5$?

$R^2 = 0.7^2 \cong 0.5$, 50% of original variation is explained

$R^2 = 0.5^2 = 0.25$, 25% of original variation is explained.

With R^2, it is easy to see that the first correlation is twice as good as second. But R^2 does not indicate direction of correlation because square is never negative.

Adjusted R^2 →
- Suppose we have two variable, Revenue of Ice cream vendor & temperature if $R^2 = 80$, this means 80% of increase in ice cream cart revenue is due to increase in temperature
- Let add another useless variable (independent), level of education of workers. R^2 increase and $R^2 = 85\%$
- One limitation of R^2 that it increases by adding independent variable to the model
- Adjusted R^2 overcome issue by adding penalty if new variable does not improve model.
- If useless variable added to model, Adjusted R^2 will decrease and if useful predictor are added to model, Adjusted R^2 will increase.